

دریافت: ۹۳/۴/۱۳

پذیرش: ۹۳/۹/۹

بازشناسی انسان در سیستم‌های نظارت ویدئویی

علی سبٹی^{۱*}، حمید حسن پور^۲

^۱ دانشجوی دکتری، دانشکده کامپیوتر و فناوری اطلاعات، دانشگاه شاهرود، شاهرود، ایران

ali.sebti@shahroodut.ac.ir

^۲ استاد، دانشکده کامپیوتر و فناوری اطلاعات، دانشگاه شاهرود، شاهرود، ایران

h.hassanpour@shahroodut.ac.ir

چکیده: بازشناسی افراد یکی از پردازش‌های مهم و پایه‌ای در سیستم‌های نظارت ویدئویی می‌باشد. دقت و کارآمدی این پردازش می‌تواند منجر به کارآمدی پردازش‌های بعدی گردد. از پردازش‌های بعدی که در سطوح بالاتر معنایی قرار می‌گیرند، می‌توان به آشکارسازی وقایع، تجزیه و تحلیل رفتاری افراد اشاره کرد. در این پردازش می‌خواهیم با در دست داشتن تصویر یا ویدئویی از یک فرد در دوربینی مشخص، تشخیص دهیم که آیا این فرد در گذشته در دوربین دیگری رؤیت شده است. انواع تغییرات ظاهری افراد از دوربینی به دوربین دیگر، مشکلات گوناگونی را در بازشناسی افراد ایجاد می‌کند. این تغییرات می‌تواند به دلیل محدودیت‌های سخت‌افزاری مانند کیفیت پایین دوربین‌ها به منظور کاهش هزینه‌ها، پاسخ‌های فرکانسی متفاوت دوربین‌ها نسبت به طیف‌های رنگی، تغییر شرایط نوری در موقعیت‌های مختلف نصب دوربین‌ها یا تغییر زاویه دید دوربین نسبت به فرد بروز یابد. در این مقاله، ضمن مروری بر روش‌های موجود در نظارت ویدئویی، دسته‌بندی جدیدی از آن‌ها براساس نحوه عملکرد ارائه خواهیم کرد. بررسی ما نشان می‌دهد که مشکل تغییر زاویه دید دوربین‌ها نسبت به یک فرد در روش‌های موجود، مورد توجه جدی قرار نگرفته و مسیر تحقیقات آتی می‌تواند در راستای ارائه مدل‌هایی با در نظرگیری این مشکل باشد.

واژه‌های کلیدی: بازشناسی افراد، ردیابی، سیستم نظارت ویدئویی.

۱. مقدمه

از فرضیاتی که در اکثر روش‌های ارائه شده مورد استفاده قرار می‌گیرد این است که ظاهر افراد از یک دوربین به دوربین دیگر بدون تغییر است [۱]. در نتیجه، ارائه روش و سیستم‌های هوشمند و خودکاری که بتواند بر مشکلات اشاره شده در بالا فائق آید از اهمیت ویژه‌ای برخوردار است. این سیستم‌ها علاوه بر اینکه باید پیچیدگی محاسباتی قابل قبولی داشته باشند، باید نسبت به تغییرات زاویه دید نسبت به دوربین‌های مختلف، شرایط نوری گوناگون، هم‌پوشانی‌های احتمالی رخ داده و مسائل مشابه مقاوم باشند. پیچیدگی دوربین‌ها نیز از فرضیات دیگر می‌باشد که انواع گوناگونی را شامل می‌گردد که عبارت‌اند از: تک دوربین در مقابل چند دوربین، دوربین‌های با میدان دید مشترک در مقابل دوربین‌های غیر همپوشان، با نظارت و تعامل کاربر در مقابل کاملاً خودکار، دوربین‌های ثابت در مقابل دوربین‌های متحرک، مجموعه باز در مقابل مجموعه بسته. منظور از مجموعه بسته^۱ این است که افراد قابل رؤیت در سیستم، از پیش مشخص می‌باشند. در حالت مجموعه باز^۲ افراد به مرور زمان رؤیت می‌گردند و از پیش مشخص نیستند.

هدف از این پژوهش بررسی مهم‌ترین و جدیدترین روش‌های ارائه شده در بازشناسی انسان می‌باشد. همچنین دسته‌بندی جدیدی از این روش‌ها ارائه خواهیم کرد. دسته‌بندی جدید نقایص دسته‌بندی‌های کنونی را نداشته و از منظر تفاوت این دسته‌بندی را انجام خواهد داد. در ادامه ویژگی‌های روش‌های مختلف را بر خواهیم شمرد و مقایسه‌ای از آن‌ها را ارائه خواهیم کرد. در انتها با برشمردن نقاط مشترک و نقایص این روش‌های پیشنهادی در چگونگی ادامه این تحقیقات در آینده که برطرف‌کننده این نقایص هستند ارائه خواهیم کرد.

در فصل دوم، این مقاله مروری بر دسته‌بندی‌ها و روش‌های گوناگون ارائه شده در بازشناسی افراد خواهیم داشت. همچنین در این فصل دسته‌بندی پیشنهادی خود را ارائه خواهیم نمود. در فصل سوم مروری بر معیارهای ارزیابی بازشناسی انسان خواهیم داشت. در فصل چهارم مجموعه داده‌ها برای ارزیابی عملکرد

سیستم‌های نظارت ویدئویی امروزه در موقعیت‌های مختلفی از جمله اماکن عمومی، به‌طور گسترده‌ای مورد استفاده قرار می‌گیرند. هدف از استقرار این سیستم‌ها مانیتور کردن موقعیت و رفتار افراد است. بازشناسی افراد یکی از فازهای اولیه و پیش‌نیاز در پردازش‌های سطح بالاتری از قبیل شناسایی و آنالیز فعالیت و رفتاری افراد، بازشناسی وقایع و رویدادها و تجزیه و تحلیل صحنه است. به عبارتی، پردازش‌های ذکر شده زمانی بهترین نرخ تشخیص را خواهند داشت که بازشناسی افراد با دقت خوبی صورت گرفته باشد.

در سیستم‌های نظارت ویدئویی، سخت‌افزارهای مورد استفاده، دوربین‌های مدار بسته‌ای می‌باشند که داده‌ها را فراهم می‌آورند. در این سیستم‌های هوشمند و خودکار، انواع پردازش‌های بینایی ماشین صورت می‌گیرد. بازشناسی افراد نیز یکی از جنبه‌های اصلی این پردازش‌هاست. براساس تحقیقات صورت گرفته، بازشناسی افراد را می‌توان چنین تعریف کرد: فرایندی است که در آن با در دست داشتن تصویر یا دنباله‌ای از تصاویر یک فرد در یک دوربین، به دنبال شناسایی مجدد همان فرد در دوربین‌های دیگر هستیم [۱]؛ برای مثال، در یک سیستم نظارتی با مشاهده یک فرد در یک دوربین، مشخص می‌گردد که این فرد قبلاً در کدام یک از دوربین‌های سیستم نظارتی مذکور رؤیت شده است؛ در صورتی که این فرد قبلاً رؤیت شده باشد، سیستم برچسبی معادل برچسب فرد پیشین را در نظر می‌گیرد و در غیر این صورت، فرد مذکور فرد جدیدی است که اکنون وارد محیط نظارتی شده و سیستم یک برچسب جدید برای آن فرد در نظر می‌گیرد. حتی زمانی که در یک دوربین هستیم و به دلیلی ردیابی یک شخص در همان دوربین مختل و گم می‌گردد، به منظور شناسایی مجدد فرد و بازیابی مجدد ردیابی، این پردازش می‌تواند سودمند باشد. تعدد افراد که فضای جست‌وجو را بزرگ کرده، کیفیت پایین دوربین‌ها، شرایط نوری متفاوت دوربین‌ها، زاویه‌های مختلف افراد در دوربین‌های مختلف شرایط بازشناسی را مشکل‌تر می‌کند [۲].

1. Close set

2. Open set

سیستم‌های نظارت ویدئویی معرفی گردیده و در فصل پنجم بحث و نتیجه‌گیری تحقیق صورت گرفته را ارائه خواهیم نمود.

۲. انواع روش‌های بازشناسی افراد

از منظر دسته‌بندی عملکردی، مقالات موجود دسته‌بندی‌های گوناگونی را ارائه نموده‌اند. این دسته‌بندی‌ها مشابهت‌هایی با یکدیگر دارند. در اغلب این دسته‌بندی‌ها روش‌های موجود به دو دسته کلی تقسیم می‌شوند: روش‌هایی که تنها از اطلاعات ظاهر استفاده می‌کنند و روش‌هایی که از اطلاعات پیکربندی دوربین‌ها استفاده می‌کنند. در دسته روش‌هایی که تنها از اطلاعاتی تصاویر استفاده می‌کنند دو زیردسته فعال و غیرفعال تعریف گردیده است. منظور از فعال روش‌هایی است که در مراحل مختلف از فرایند یادگیری استفاده می‌کنند و روش‌های غیرفعال که از فرایند یادگیری استفاده نمی‌کنند. مشابه دسته‌بندی یادشده در مقالات [۱]، [۲] می‌باشد که شکل ۱ این دسته‌بندی را در قالب یک نمودار درختی نشان می‌دهد.

به‌طور کلی، روش‌های بازشناسی سعی در ارائه مدلی برای یک فرد رؤیت شده در سیستم دارند. در مرحله شناسایی، تصویر فرد مورد آزمون با مدل‌های ساخته‌شده مقایسه شده و تصمیم‌گیری به منظور برچسب‌گذاری انجام می‌شود. همچنین، مهم‌ترین فاز در مدل‌سازی تخمین پارامترهای مدل می‌باشد. در بیشتر روش‌ها تخمین پارامتر نیز در فرایند یادگیری صورت می‌پذیرد.



شکل (۱): دسته‌بندی روش‌های بازشناسی افراد در سیستم‌های نظارتی

به‌طور خلاصه، می‌توان گفت که هر جا مدلی داریم، یادگیری نیز وجود دارد؛ برای مثال، با توجه به شکل ۱ برخی از روش‌ها که در زیر دسته «مدل‌های برپایه ظاهر» قرار دارند نیز در مرحله تخمین پارامتر، به‌نوعی از فرایند یادگیری بهره می‌برند. یا در زیر دسته «استخراج موقعیت نسبی قرارگیری دوربین‌ها» برای موقعیت نسبی دوربین‌ها مدلی احتمالاتی در نظر گرفته و در فرایند یادگیری پارامترهای مدل تخمین زده می‌شود. به نظر می‌آید که مقصود نویسنده از یادگیری، استفاده از مجموعه آموزشی بزرگ است، درحالی‌که از این منظر، در روش‌های غیرفعال فاز یادگیری مشهودی نداریم.

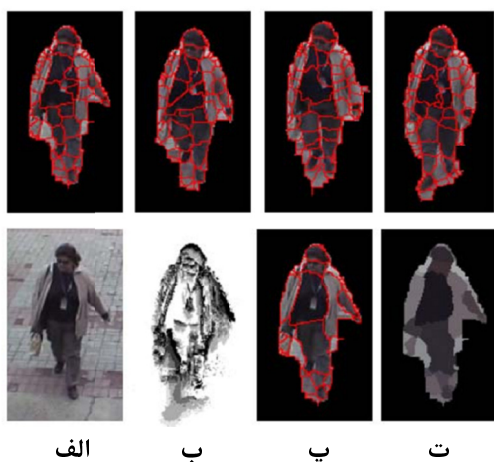
براساس ایراد مطرح‌شده ما نیز سعی داریم دسته‌بندی جدیدی برای روش‌های بازشناسی افراد ارائه کنیم. شکل ۲ این دسته‌بندی را در قالب نمودار نشان می‌دهد. در دسته‌بندی پیشنهادی، تمرکز بر روی فضای جست‌وجوست. زمانی که فرد جدیدی وارد سیستم می‌شود، سیستم نظارتی با فضای جست‌وجویی از افراد رؤیت شده تاکنون روبه‌روست. گروهی از روش‌ها سعی در محدود و کوچک کردن این فضا دارند و گروهی سعی در ارائه مدل بهتری برای شخص در سیستم می‌باشند. در این گروه، مدلی بهتر است که از بین تصاویر موجود و رؤیت‌شده، کاندیدهای معرفی کند که تطبیق بهتری با تصویر درخواستی را داشته باشد.



شکل (۲): دسته‌بندی پیشنهادی روش‌های بازشناسی

در ادامه، مروری بر مهم‌ترین روش‌های ارائه‌شده تاکنون خواهیم داشت. دسته‌بندی این روش‌ها در قالب دسته‌بندی پیشنهادی شکل ۲ تنظیم شده است.

قطعه‌بندی نواحی به‌هم‌پیوسته $R = \{r_i^t\}$ است که در آن i اندیس ناحیه در تصویر t ام می‌باشد. بر این اساس، گراف $G = \{V, E\}$ با مجموعه رئوس $V = \{v_i^t\}$ و مجموعه یال‌های $E = \{e_{i,i'}^{t,t'}\}$ تعریف می‌شود. در این تعریف، رئوس گراف (v_i^t) متناظر نواحی r_i^t و یال‌ها $(e_{i,i'}^{t,t'})$ نیز دو راس v_i^t و $v_{i'}^{t'}$ را به‌هم متصل می‌کند. پس از ساخت چنین گرافی از الگوریتم‌های پارتیشن‌بندی یا خوشه‌بندی گراف استفاده شده و بخش‌بندی روی نواحی مذکور صورت می‌گیرد. شکل ۳ بخش‌بندی تصویر را در طول زمان براساس روش مقاله [۳] نشان می‌دهد. در این بازنمایی گرافی، وزن لبه‌ها براساس فاصله رنگی نواحی به‌دست آمده از مرحله قطعه‌بندی، در دو راستای زمان و مکان محاسبه می‌شود.



شکل (۳): ردیف اول نتیجه قطعه‌بندی حوزه آبریز در ناحیه پیش‌زمینه؛ ردیف دوم: قسمت ب عکس فرکانسی، قسمت پ نتیجه خوشه‌بندی، قسمت ت نمایش هر خوشه با یک رنگ (رنگ میانه خوشه)

همچنین در این مقاله، سعی در انطباق مدلی به تصویر انسان و لبه‌های بااهمیت به‌دست آمده در قسمت بخش‌بندی می‌باشد. این انطباق از طریق بهینه‌سازی یک تابع هزینه به کمک برنامه‌نویسی پویا انجام شده است. نتایج آزمایشات برتری استفاده از مدل را نشان می‌دهد. مدل مذکور یک مدل مثلثی به صورت شکل ۴ می‌باشد.

۱.۲. مدل‌هایی با تمرکز بر استخراج ویژگی ظاهری

در این گروه از روش‌ها، هدف استخراج ویژگی برای یک فرد است که یا دارای خاصیت متمایزکنندگی قوی نسبت به افراد دیگر باشد یا خاصیت توصیف‌کنندگی خوبی داشته باشد. در واقع، در این گروه به دنبال ویژگی‌ای هستیم که با در دست داشتن تصویر آزمون، بتواند افراد را در مجموعه گالری طوری مرتب کند که تصویر فرد مورد جست‌وجو در رتبه بالاتری قرار گیرد. این ویژگی‌ها باید نسبت به انواع مشکلات ذکرشده در بخش مقدمه مقاوم باشند.

۱.۱.۲. مدل‌هایی با خاصیت توصیف‌کنندگی قوی

در این دسته، هدف ساخت مدل‌هایی از انسان است که بازگوکننده بهتری از یک انسان باشد. در عین حال سعی می‌شود مدلی ارائه گردد که نسبت به انواع تغییرات، مقاومت خوبی از خود نشان دهد. در این باره تحقیقات و مقالات گوناگونی ارائه شده که در ادامه، به برخی از مهم‌ترین این روش‌ها اشاره خواهیم کرد.

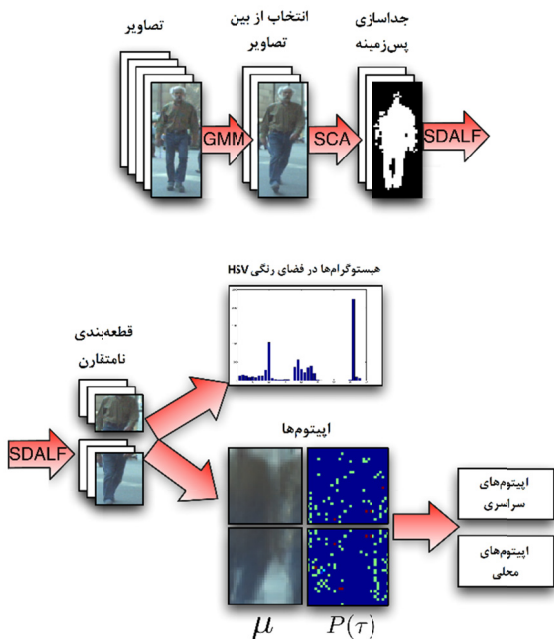
• مدل ظاهری مکانی-زمانی

یکی از مشکلات مسئله بازشناسی انسان این است که در طول حرکت یا براساس موقعیتی که فرد در آن قرار دارد، ممکن است لباس فرد دستخوش حرکتی شود. به عبارتی، چین‌وچروک‌هایی ممکن است بر روی لباس فرد ایجاد شود که کیفیت بازشناسی را تحت تأثیر قرار دهد. در مقاله [۳] سعی شده بر این مشکل غلبه گردد. ایده اصلی مطرح شده در این مقاله، معرفی الگوریتم بخش‌بندی جدیدی تحت عنوان بخش‌بندی مکانی-زمانی است.

در این مقاله، ابتدا قاب‌های مربوط به یک فرد، به دو بخش پیش‌زمینه و پس‌زمینه تقسیم می‌شود. این تقسیم‌بندی براساس میزان فراوانی مقدار رنگی پیکسل‌ها در یک پنجره، زمانی خاص تعیین می‌شود. سپس ناحیه پیش‌زمینه هر یک از تصاویر پنجره زمانی مذکور، قطعه‌بندی حوزه آبریز^۱ می‌گردد. نتیجه این

1. Watershed algorithm

اول، توصیفگر هیستوگرام بر روی مدل رنگی HSV^۳ می‌باشد. قسمت دوم توصیفگر، از آنالیز جوهره استفاده می‌کند. آنالیز جوهره تجمیعی از چند تصویر در قالب یک مدل بازگوکننده فراهم می‌آورد؛ به طوری که این مدل دربرگیرنده خصوصیات بافتی، شکلی و ظاهری است. این آنالیز یکبار به صورت کلی و بار دیگر به صورت محلی، روی تصاویر حاصل از خوشه‌بندی اعمال می‌شود. شکل ۵ خلاصه‌ای از این الگوریتم را نشان می‌دهد.

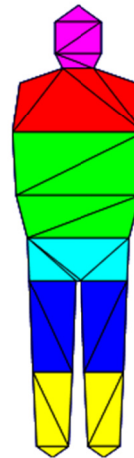


شکل (۵): خلاصه‌ای از توصیفگر ترکیبی ارائه شده در [۴]

بدین ترتیب، برای محاسبه فاصله دو مجموعه از تصاویر از ترکیب وزن‌داری از آنالیز جوهره در حالت محلی و سراسری و اطلاعات هیستوگرام استفاده شده است. در نهایت، نسخه ترکیبی این روش با قطعه‌بندی غیر قرینه‌ای ارائه شده در [۶] منجر به بازشناسی با نرخ دقت بیشتر شده است.

• **تجمیع ویژگی‌های محلی بر پایه قرینگی**

در مقاله [۶]، از بازنمایی خاصی برای انسان استفاده گردیده است. در این بازنمایی ابتدا به کمک یک روش ابتکاری ناحیه بدن



شکل (۴): تصویر راست مدل مثلی، تصویر چپ: دو نمونه انطباق به کمک لبه‌های بااهمیت و قسمت ناحیه پیش‌زمینه به دست آمده از الگوریتم بخش‌بندی مبتنی بر پارتیشن‌بندی گراف

- بازشناسی چندتصویره انسان به کمک توصیفگر HPE در مقاله [۴] از یک روش ترکیبی برای ارائه توصیفی از چندین تصویر به طور توأمان برای انسان استفاده شده است. قسمت اصلی این الگوریتم از دو بخش هیستوگرام و جوهره^۱ تشکیل می‌شود (HPE^۱). در اولین مرحله نواحی پیش‌زمینه به کمک الگوریتم ارائه شده در [۵] جداسازی می‌گردد. قسمت

1. Epitome
2. Histogram Plus Epitome

3. Hue. Saturation. Value Color Space

ث ت پ ب الف



شکل (۶): ستون الف: تصویر اصلی، ستون ب: استخراج نواحی تنه و پاها و محورهای قرینگی، ستون پ: تصویر back-projection هیستوگرام وزن دار، ستون ت: توصیفگر MSCR و ستون ث: تکه RHSP

• روش‌های مبتنی بر توصیفگر کوواریانس

توصیفگر کوواریانس به نوعی خصوصیات بافتی تصویر را بازگو می‌کند. از خصوصیات مطلوب این توصیفگر مقاومت آن نسبت به چرخش، تغییرات نوری و اندازه می‌باشد؛ لذا می‌توان از آن در کاربرد بازشناسی نیز استفاده کرد. ایده اصلی در این توصیفگر، محاسبه ماتریس کوواریانس تعدادی بردار ویژگی هم‌بعد برای یک ناحیه از تصویر است. این ویژگی‌ها می‌تواند شامل موقعیت، کانال‌های رنگی و پاسخ‌های ناحیه از تصویر به انواع فیلترها باشد. در واقع، توصیفگر کوواریانس همبستگی بین چندین نوع ویژگی را محاسبه می‌کند. این ماتریس می‌تواند خصوصیات ساختاری و بافتی آن ناحیه از تصویر را بازنمایی کند. نحوه محاسبه این توصیفگر بدین ترتیب است که فرض کنید R_1 ، ناحیه‌ای از تصویر باشد که می‌خواهیم این توصیفگر را برای آن محاسبه کنیم. برای هر پیکسلی که درون این ناحیه قرار می‌گیرد، ویژگی‌های مورد نظر از قبیل رنگ، گرادیان و پاسخ‌های فیلتری محاسبه می‌شود. بدین ترتیب، برای هر پیکسل برداری d بعدی از ویژگی‌ها را خواهیم داشت؛ برای مثال، در کاربرد بازشناسی انسان، می‌توان سه کانال رنگی R و G و B و خصوصیت y را ویژگی در نظر گرفت. دلیل انتخاب خصوصیت y این است که تصویر انسان در

فرد به سه قسمت سر، تنه و پاها تقسیم‌بندی می‌شود. با تغییر اندکی از همان روش، برای استخراج محور عمودی در هر بخش که فرض قرینگی برای آن صادق است، استفاده شده است. برای هر یک از این نواحی سه ویژگی استخراج می‌گردد. اولین ویژگی، هیستوگرام وزن دار در مدل رنگی HSV می‌باشد. طریقه وزن دار کردن آن نیز بدین ترتیب است که هرچه یک پیکسل در تصویر فرد از محور عمودی قرینه دور باشد وزن کمتری در هیستوگرام خواهد گرفت. دلیل این وزن دهی در این موضوع نهفته که هرچه از محور عمودی دور می‌شویم احتمال تعلق پیکسل به پس زمینه بیشتر می‌گردد. دومین ویژگی مورد استفاده توصیفگر $MSCR^1$ می‌باشد. توصیفگر MSCR در واقع خوشه‌بندی بر روی پیکسل‌های هر ناحیه از تصویر شخص انجام می‌دهد. در حین خوشه‌بندی این نواحی به طوری انتخاب می‌شوند که به لحاظ رنگی فاصله بیشتری نسبت به هم داشته باشند. هر خوشه با نه بعد نمایش داده می‌شود که شامل مساحت، مرکز خوشه، میانگین رنگی RGB و ماتریس تکانه^۲ دوم می‌باشد. در حالت چند تصویر خوشه‌بندی دیگری بر روی این نواحی برای چند تصویر انجام می‌شود. سومین ویژگی مورد استفاده نواحی با خصوصیت ساختاری بالا و مکرر ($RHSP^3$) می‌باشد. ایده استخراج چنین نواحی برگرفته از آنالیز جوهره می‌باشد. در واقع در این روش ما به دنبال نواحی هستیم که دارای مقدار آنتروپی بالا بوده و در عین حال تحت تبدیل‌های مختلف نیز بیشتر در تصویر رؤیت شوند. به عبارتی نواحی که تحت تبدیلاتی از قبیل چرخش، اندازه، بیشتر در تصویر دیده می‌شوند نسبت به تغییرات زاویه‌ای انسان نیز مقاوم می‌باشند. شکل ۶ الگوریتم $SDALF^4$ و سه ویژگی استخراجی را نشان می‌دهد.

1. Maximally Stable Color Regions
2. Momentum
3. Recurrent High-Structured Patches
4. Symmetry-Driven Accumulation of Local Features

راستای X به دلیل فرینگی متمایزکننده نیست.

با فرض اینکه ابعاد تصویر W در H باشد، مجموعه بردارهای ویژگی $\{f_i\}_{i=1 \dots W*H}$ به صورت زیر تعریف می گردد [۹]:

$$f_i = (y, R(x, y), G(x, y), B(x, y))^T \quad (1)$$

بر اساس این بردارهای ویژگی، ماتریس کوواریانس به صورت زیر محاسبه می گردد:

$$Cov_{R_1} = \frac{1}{WH} \sum_{i=1}^{WH} (f_i - \mu_{R_1})(f_i - \mu_{R_1})^T$$

$$\mu_{R_1} = \frac{1}{WH} \sum_{i=1}^{WH} f_i \quad (2)$$

که در آن، بردار میانگین بردار ویژگی هاست. بدین ترتیب، ماتریس کوواریانس برای هر ناحیه ای از تصویر با هر اندازه ای، ماتریسی مربعی و قرینه و مثبت معین با ابعاد $d \times d$ می باشد. نکته ای که در کار کردن با ماتریس کوواریانس باید بدان توجه کرد، این است که چنین ماتریسی در فضای اقلیدسی قرار ندارد. به عبارتی، برای محاسبه شباهت دو ماتریس کوواریانس نمی توان از فاصله اقلیدسی استفاده کرد. ماتریس های کوواریانس روی رویه ای^۱ ریمانی قرار گرفته اند. در فضای ریمانی^۲، فاصله یا عدم شباهت دو ماتریس کوواریانس C_i و C_j از رابطه زیر محاسبه می شود:

$$\rho(C_i, C_j) = \sqrt{\sum_{k=1}^d \ln^2 \lambda_k(C_i, C_j)} \quad (3)$$

که در آن، $\lambda_k(C_i, C_j)_{k=1 \dots d}$ مقادیر ویژه تعمیم یافته دو ماتریس C_i و C_j است که از رابطه زیر محاسبه می شود:

$$\lambda_k C_i x_k - C_j x_k = 0 \quad k = 1 \dots d \quad (4)$$

در مقاله [۷] بازشناسی را در سیستمی چندتصویره انجام می دهد. بدین ترتیب که برای یک فرد دنباله ای از تصاویر

داریم. برای هریک از این تصاویر، توصیف گر کوواریانس محاسبه می شود. سپس گرافی تشکیل می دهیم که رئوس آن ماتریس های کوواریانس و یال ها میزان فاصله ماتریس ها از هم می باشد که در بالا ذکر شد. سپس بر روی گراف به دست آمده خوشه بندی طیفی صورت گرفته و در نتیجه، برای هر دنباله k میانگین محاسبه می شود و در نهایت، در مرحله بازشناسی این k میانگین ها با هم مقایسه و شباهت تعیین می گردد.

در مقاله [۸] به سه کانال رنگی HSV تجزیه کرده و سپس به کمک فیلتر گابور در ۸ باند و در هر باند دو سایز مختلف روی هر کانال رنگی اعمال می شود. در این مقاله، برای بالا بردن سرعت محاسبات به جای محاسبه ماتریس کوواریانس برای کل تصویر، با تقسیم بندی تصویر به نواحی کوچک تر و محاسبه ماتریس مذکور برای هر ناحیه، سرعت انجام عملیات را بالا برده است. در قسمتی از الگوریتم ارائه شده، فاصله (رابطه ۱) نواحی متناظر برای دو سایز از پاسخ فیلتر گابور^۳ محاسبه و برداری از این مقادیر برای بازنمایی فرد مورد استفاده قرار می گیرد.

در مقاله [۹] فاصله از مرکز تصویر به عنوان یک ویژگی کلیدی در محاسبه ماتریس کوواریانس لحاظ شده است. فاصله از مرکز ارتباط با نواحی پراهمیت تصویر را حفظ می کند. در این مقاله، تعدادی ویژگی در نظر گرفته شده، سپس نواحی مختلفی در تصویر یک فرد، به صورت تصادفی در اندازه های مختلف انتخاب می شود. سپس از بین نواحی که به صورت تصادفی تعریف شده با استفاده از روش CFS^۴ ویژگی های بااهمیت استخراج می گردد؛ به طوری که برای داده های یک کلاس فاصله کمتر و برای داده های غیر یک کلاس، فاصله بیشتر گردد. پس از استخراج ویژگی ها می توانیم جفت تصاویر افراد را با هم مقایسه کنیم. در مقاله [۱۰] نیز از فیلتر گابور و LBP^۵ برای بردار ویژگی استفاده کرده و همچنین، ناحیه بدن فرد را به چهار قسمت تقسیم بندی نموده و روی هر ناحیه، توصیف گر کوواریانس را

3. Gabor filter

4. Correlation based Feature Selection

5. Local Binary Pattern

1. Manifold

2. Riemannian space

سپس برای هر موقعیت از سلول‌ها با استفاده از روش کارچر که با کمک الگوریتمی تکراری براساس روش نزول در جهت کاهش گرادیان، میانگین n ماتریس کوواریانس در دنباله مذکور براساس رابطه زیر محاسبه می‌شود.

$$\mu_{t+1} = \exp_{\mu_t} \left[\frac{1}{N} \sum_{i=1}^N \exp_{\mu_t}(C_i) \right] \quad (6)$$

در این رابطه، \exp_{μ_t} و \exp_{μ_t} عملگرهای مخصوصی در رویه ریمانی می‌باشند. خروجی این مرحله MRCG نامیده می‌شود. بدین ترتیب برای n فرد در سیستم مجموعه n امضا $\mathcal{S}^c = \{\mathcal{s}_i^c\}_{i=1}^n$ به صورت زیر خواهیم داشت (m تعداد سلول‌ها در شبکه است):

$$\mathcal{s}_i^c = \{\mu_{i,1}^c, \mu_{i,2}^c, \dots, \mu_{i,m}^c\} \quad (7)$$

در قدم بعدی، این مقاله برای هر فرد سعی دارد میزان اهمیتی برای هر سلول درون شبکه آن تعریف کند. بدین منظور، از تعریف واریانس استفاده می‌کند. بدین ترتیب که سلولی که در مجموع تفاوت بیشتری با سلول متناظر آن در بین کل افراد داشته باشد، برای فرد مذکور سلول بااهمیتی است. در نتیجه با این روش، سلول‌های پس‌زمینه به دلیل اینکه واریانس زیادی ندارند، وزن کمتری گرفته و همچنین تمرکز روی تفاوت‌های فرد با بقیه افراد است. بر این اساس، واریانس هر سلول از رابطه زیر محاسبه شده است:

$$\sigma_{i,j}^c = \frac{1}{n-1} \sum_{k=1, k \neq i}^n \rho^2(\mu_{i,j}^c, \mu_{k,j}^c) \quad (8)$$

در نهایت شباهت دو فرد به صورت زیر محاسبه می‌شود:

$$S(\mathcal{s}_A, \mathcal{s}_B) = \sum_{i \in K} \frac{\sigma_{A,i} + \sigma_{B,i}}{\rho(\mu_{A,i}, \mu_{B,i})} / \|K\| \quad (9)$$

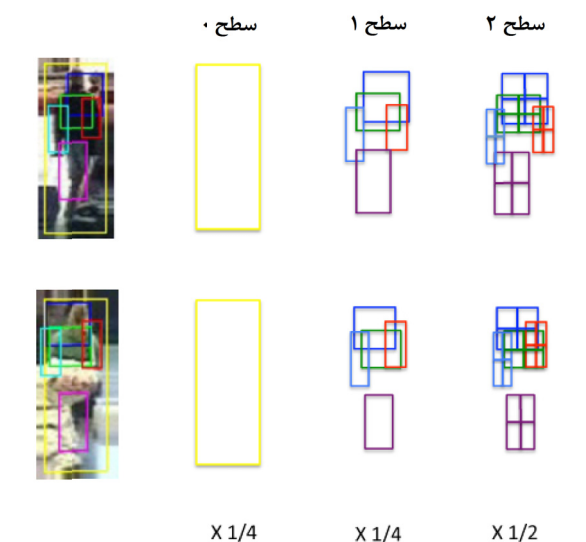
• بازشناسی بر پایه هیستوگرام رنگی

هیستوگرام یکی از ویژگی‌های مناسب به منظور بازنمایی یک تصویر است. دلیل استفاده گسترده از این ابزار، سرعت پردازش‌هایی است که روی هیستوگرام انجام می‌گیرد؛ از قبیل محاسبه هیستوگرام، مقایسه دو هیستوگرام. از مشکلات اصلی

براساس ویژگی‌های یادشده محاسبه می‌کند. در مقاله [۱۱] از یک ساختار هرمی استفاده شده است. بدین ترتیب که ابتدا به کمک توصیف‌گر HOG و کلاس‌بندی که به منظور استخراج ناحیه‌ای از تصویر که انسان در آن واقع شده و همچنین اعضای بدن آموزش دیده است، ناحیه فرد و نواحی اعضای بدن را تشخیص می‌دهیم. شکل ۷ ساختار هرمی سه‌سطحی معرفی شده در این مقاله را نشان می‌دهد. همان‌طور که در شکل مشاهده می‌شود، سطح اول مربوط به ناحیه کل بدن فرد است. سطح بعدی مربوط به اعضای بدن فرد و سطح سوم مربوط به زیر ناحیه اعضای بدن فرد می‌باشد. فرایند تطبیق در سیستم‌هایی به شکل چند سطحی و هرمی بدین صورت است که برای سطوح جزئی‌تر، وزن بیشتری اختصاص داده می‌شود و برای سطوح کلی‌تر، وزن کمتر. بردار ویژگی‌های استفاده شده در این مقاله نیز به صورت زیر است:

$$[x, y, R_{xy}, G_{xy}, B_{xy}, \|\nabla_{xy}^R\|, \theta_{xy}^R, \|\nabla_{xy}^G\|, \theta_{xy}^G, \|\nabla_{xy}^B\|, \theta_{xy}^B] \quad (5)$$

در مقاله [۱۲]، ابتدا شبکه‌ای متراکم (با سلول‌های هم‌پوشان جزئی) روی دنباله‌ای از تصاویر یک فرد ایجاد می‌کنیم. سپس برای هر سلول از شبکه یادشده، توصیف‌گر کوواریانس را حساب می‌کنیم.



شکل (۷): ساختار هرمی مورد استفاده در مقاله [۱۱]

² Mean Riemannian Covariance Grid

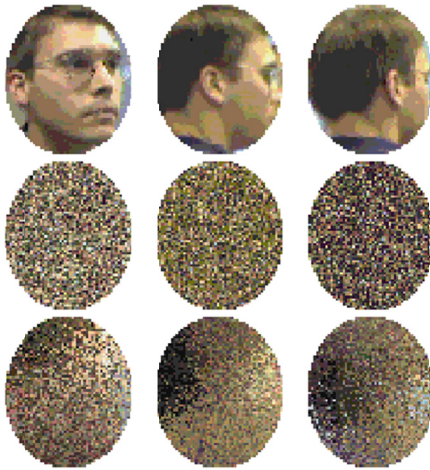
انبارک مربوط واقع می‌شوند. با این تعریف جدید، شباهت دو هیستوگرام مکانی نیز به صورت زیر تعریف می‌شود:

$$\rho(h, h') = \sum_{b=1}^B \psi_b \rho_n(n_b, n'_b)$$

$$\psi_b = \eta \cdot \exp \left\{ -\frac{1}{2} (\mu_b - \mu'_b) \Sigma_b^{-1} (\mu_b - \mu'_b) \right\}$$

$$\Sigma_b^{-1} = \Sigma_b^{-1} + (\Sigma'_b)^{-1} \quad (12)$$

فرض کنید هیستوگرام تصاویر سطر اول شکل ۸ را داشته باشیم. اگر بخواهیم از روی این هیستوگرام تصویری ایجاد کنیم، به صورت تصادفی در تمام ناحیه تصویر نقاطی با توزیع مشخص شده در هیستوگرام ایجاد می‌کنیم که تصویری مشابه سطر دوم شکل ۸ تشکیل می‌شود. براساس هیستوگرام مکانی مرتبه دوم تصویر به دست آمده، شباهت بیشتری با تصویر اولیه دارد. به عبارتی خصوصیات مکانی بهتر حفظ شده‌اند. در مقاله [۱۶] با استفاده از بخش بندی تصویر اطلاعات مکانی را به هیستوگرام می‌افزاید. در مرحله بعد، برای هر ناحیه، هیستوگرام به صورت مجزا محاسبه می‌شود.



شکل (۸): مقایسه بازسازی یک تصویر از روی هیستوگرام مکانی و هیستوگرام استاندارد

در این مقاله، ابتدا میان تنه تصویر فرد استخراج شده و چهار هیستوگرام به صورت شعاعی به مرکزیت مرکز نیم‌تنه

هیستوگرام در نسخه استاندارد، عدم حفظ شدن اطلاعات مکانی در تصویر است.

در مقاله [۱۳] تعریف جدیدی از هیستوگرام رنگی ایجاد شده است که به آن، هیستوگرام رنگی احتمالاتی گفته می‌شود. روش کار بدین ترتیب است که برای همه رنگ‌های موجود ۱۱ رنگ مرسوم را که در فرهنگ‌های مختلف وجود دارد، به عنوان رنگ‌های اصلی انتخاب می‌کنیم. سپس با کمک مجموعه داده‌های پرچسب‌خورده از تصاویر مختلف حاوی رنگ در شرایط محیطی گوناگون، با استفاده از الگوریتم خوشه‌بندی k نزدیک‌ترین همسایه فازی فضای رنگی اولیه LAB به ۱۱ رنگ به صورت فازی تقسیم‌بندی می‌شود. در نتیجه با این تعریف، یک پیکسل با توجه به رنگ آن می‌تواند بیش از یک انبارک از هیستوگرام را تغییر دهد. ایده اصلی مقاله [۱۴] نیز استفاده از پنج ویژگی R, G, B, x و y در هیستوگرام است. بدین ترتیب که هیستوگرامی ۵ بعدی براساس پنج ویژگی فوق تشکیل داده و سپس برای هر پیکسل با توجه به موقعیت و رنگ، درجه عضویتی به یکی از این انبارک‌ها قائل شده است. بدین ترتیب، همانند روش قبلی یک پیکسل بر روی بیش از یک انبارک از هیستوگرام تأثیر خواهد گذاشت.

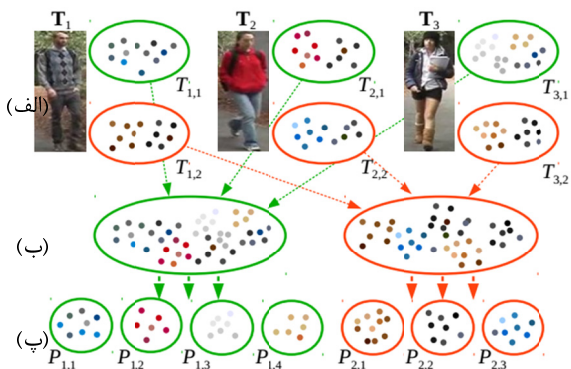
در مقاله [۱۵] نویسنده تعمیمی از هیستوگرام استاندارد را ارائه می‌کند. بر طبق تعریف مقاله، هیستوگرام استاندارد یک Spatiogram مرتبه صفر است. تصویر I را نگاشتی از فضای X به فضای v به صورت زیر در نظر می‌گیریم:

$$I: x \rightarrow v \quad x = [x, y]^T \quad (10)$$

بدین ترتیب، هیستوگرام مکانی مرتبه دوم به صورت زیر تعریف می‌شود:

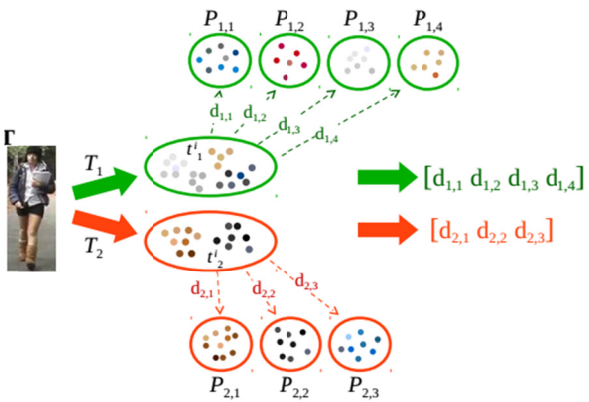
$$h_l^2(b) = \langle n_b, \mu_b, \Sigma_b \rangle b = 1, \dots, B \quad (11)$$

که در آن، n_b تعداد پیکسل‌های بین b و μ_b و Σ_b به ترتیب، میانگین مکانی و کوواریانس مکانی پیکسل‌هایی هستند که در



شکل (۱۰): الف. بخش‌بندی تصویر فرد به دو ناحیه و استخراج مؤلفه‌های ویژگی؛ ب. ادغام همه مؤلفه‌های مرتبط با یک بخش؛ پ. خوشه‌بندی هر بخش و استخراج نماینده‌ها [۳۸]

در هنگام بازشناسی یک فرد جدید، مؤلفه‌های هر بخش را با نماینده‌ها مقایسه کرده و فاصله هر بخش با تک‌تک نماینده‌ها محاسبه و در برداری ذخیره می‌شود. بردار به‌دست‌آمده در مرحله بعد با بردار به‌دست‌آمده برای مجموعه گالری مقایسه شده و بر این اساس، اعضای گالری مرتب می‌شوند. شکل ۱۱ مرحله بازشناسی را نشان می‌دهد.

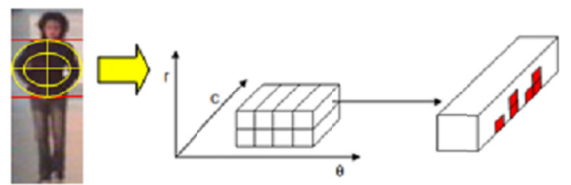


شکل (۱۱): مرحله بازشناسی در بستر عدم شباهت [۳۸]

روش‌های یادگیری توصیف‌گر

در این دسته از روش‌ها هدف، آموزش متمایزکننده‌ترین مجموعه از ویژگی‌هاست؛ برای مثال، به کمک آموزش، زیرمجموعه‌ای از ویژگی‌ها که خاصیت متمایزکنندگی بهتری دارند انتخاب می‌شوند. حالت دیگر به جای انتخاب ویژگی‌ها، وزن‌دهی به ویژگی‌هاست که بیشترین تمایز را بین نمونه‌های مختلف افراد

ایجاد می‌شود. شکل ۹ هشت هیستوگرام ذکرشده را براساس شعاع، زاویه در دایره و رنگ نشان می‌دهد.



شکل (۹): تقسیم‌بندی ناحیه تنه براساس روش ارائه‌شده در [۱۶]

در نهایت، این مجموعه از هیستوگرام‌ها به کمک معیار فاصله باتچریا^۱ در مرحله بازشناسی مورد مقایسه قرار می‌گیرد.

۲.۱.۲. مدل‌هایی با خاصیت متمایزکنندگی قوی

در این دسته، هدف، ساخت مدل‌هایی از انسان است که تمایز بهتری بین افراد مختلف ایجاد گردد. به عبارتی در این مدل‌ها تمرکز بر روی تفاوت‌های یک نمونه با بقیه نمونه‌ها می‌باشد.

• بازشناسی سریع براساس بازنمایی عدم شباهت

در مقاله [۱۷]، مسئله بازشناسی در قالب برداری از مؤلفه‌های عدم شباهت بازنمایی می‌شود. هدف اصلی از این بازنمایی پایین آوردن پیچیدگی محاسباتی در مرحله بازشناسی می‌باشد. اما دقت تشخیص در روش ارائه‌شده در حد قابل قبولی است، اما پیچیدگی محاسباتی پایینی دارد. در این معماری، ابتدا باید ناحیه بدن فرد به چند بخش تقسیم‌بندی شود. سپس برای هر بخش بردار ویژگی ایجاد گردد. در قدم بعدی، برای مجموعه‌ای از تصاویر افراد برای هر بخش به تفکیک خوشه‌بندی انجام می‌گیرد؛ برای مثال، برای قسمت تنه این خوشه‌بندی منجر به تولید ۴ خوشه می‌شود. برای پایین‌تنه نیز فرض کنید سه خوشه تولید گردد. شکل ۱۰ تعیین نماینده‌ها را نشان می‌دهد.

1. Bhattacharyya distance

کمک بوسستینگ تطبیقی، از بین ویژگی‌ها و ضرایب dcd زیرمجموعه‌ای از ویژگی‌ها که هر فرد را از بقیه بهتر متمایز می‌کند، انتخاب می‌شود [۲۱].

۲.۲. مدل‌هایی برای کوچک کردن فضای جست‌وجو

در این دسته از مدل‌ها تمرکز روی استخراج ویژگی نیست، بلکه هدف محدود کردن فضای جست‌وجوست. این محدودسازی می‌تواند به کمک استخراج ارتباط بین زوج دوربین‌ها حاصل شود. این ارتباط می‌تواند همبستگی مکانی نمونه‌های رؤیت‌شده در دوربین‌ها باشد یا تابع نگاشتی باشد که تحت آن نگاشت، خصوصیات ظاهری افراد در انتقال از یک دوربین به دوربین دیگر تغییر کند. استفاده از این ارتباط در کنار ویژگی‌های استخراج‌شده، منجر به محدود شدن فضای جست‌وجو به نمونه‌های مرتبط‌تر در مجموعه گالری می‌شود.

۱.۲.۲. مدل‌های استخراج‌کننده تابع نگاشت کلی

در این گروه، رابطه بین دو دوربین به صورت مستقیم یا غیرمستقیم مشخص می‌شود. در واقع این رابطه، نگاشتی از یک تصویر در دوربین اول به تصویر همان فرد در دوربین دوم ایجاد می‌کند. این تابع نگاشت می‌تواند در قالب یک مسئله کلاس‌بندی باشد. بدین ترتیب که برای جفت داده‌ها برای دو دوربین از یک فرد، مجموعه داده‌های آموزشی مثبت و جفت حالت‌هایی از دو فرد در دو دوربین که مربوط به یک شخص نمی‌شوند، مجموعه آموزشی منفی ساخته می‌شود. در مقاله [۲۲] از این رویکرد استفاده شده و برای مجموعه ویژگی‌ها از هیستوگرام مدل رنگی HSV برای پنج ناحیه افقی در تصویر انسان استفاده گردیده است. بدین ترتیب کلاس‌بندی که این چنین آموزش داده می‌شود، به نوعی رابطه تابع انتقال بین یک جفت دوربین را مدل کرده و یاد می‌گیرد.

• روش‌های یادگیری معیار فاصله

در این دسته از روش‌ها به جای تمرکز روی انتخاب ویژگی مناسب، تمرکز روی یادگیری معیار فاصله‌ای است که دقت تطبیق را بدون در نظر گرفتن نحوه بازنمایی ظاهری حداکثر کند. در

ایجاد کند. مقاله [۱۸] به کمک الگوریتم بوسستینگ تطبیقی، ترکیبی وزن‌دار از تعدادی ویژگی به منظور تعیین شباهت بین دو فرد استفاده شده است. الگوریتم بوسستینگ تطبیقی با ترکیب مجموعه‌ای از ویژگی‌ها که قدرت کلاس‌بندی ضعیفی دارند، کلاس‌بندی قوی ایجاد می‌کند. در هر مرحله، براساس داده‌هایی که به غلط کلاس‌بندی می‌شوند، بهترین ویژگی که کلاس‌بندی بهتری انجام می‌دهد، انتخاب شده و براساس میزان خطای کلاس‌بندی وزن‌دهی می‌شود. از ترکیب این ویژگی‌های ضعیف کلاس‌بندی قوی ایجاد می‌گردد. ویژگی‌های مورد استفاده در این مقاله، عبارت‌اند از کانال‌های مدل‌های رنگی RGB, YCbCr و HSV و پاسخ فیلترهای بافتی نظیر اشمیت و گابور. ویژگی‌های انتخابی به ناحیه در تصویر نیز وابسته‌اند.

در مقاله [۱۹] دو خاصیت توصیف‌گری و متمایزکنندگی به صورت ترکیبی مورد استفاده قرار گرفته است. بدین ترتیب که به کمک توصیف‌گر کوواریانس روی مجموعه‌ای از ویژگی‌ها، براساس درخواست کاربر ۵۰ مورد ابتدایی را انتخاب و به کاربر جهت انتخاب نمایش می‌دهد. اگر فرد مذکور در ۵۰ نفر ابتدایی نبود، از لیست مرتب‌شده افراد، تعدادی از ابتدای لیست را به عنوان مورد مثبت و تعدادی از انتهای لیست را به عنوان مورد منفی برچسب می‌زند. در ادامه به کمک الگوریتم بوسستینگ، ۲۰ ویژگی متمایزکننده برای فرد استخراج‌شده و در نهایت، لیست مجدداً به وسیله مقدار قابلیت اعتماد^۱ کلاس‌بند مرتب می‌شود. در مقاله [۲۰] نیز برای دنباله‌ای از تصاویر یک فرد، توصیف‌گر کوواریانس را برای نواحی مختلف تصویر محاسبه می‌کند. سپس برای قاب‌های متوالی مقدار میانگین ماتریس کوواریانس‌ها را برای هر ناحیه محاسبه می‌کند. در قدم بعدی به کمک دو روش، تکه‌هایی را که برای یک فرد خاص، قابلیت متمایزکنندگی بیشتری دارد، مشخص می‌کند که روش اول به کمک تعیین قابلیت اعتماد یک تکه بر پایه تعریف جدیدی از واریانس در فضای ریمانی است و دوم به کمک الگوریتم بوسستینگ تکه‌های متمایزکننده‌تر را انتخاب می‌کند. همچنین، در مقاله‌ای دیگر به

فاصله نمونه‌های غیر مرتبط قسمتی از رابطه بهینه‌سازی نبوده و تنها قیدی برای مسئله‌اند.

در مقاله‌ای دیگر [۲۵] به نوعی از فاصله بین نمونه‌ها در فرایند یادگیری استفاده کرده است. در این مقاله سعی شده از داده‌های برچسب‌نخورده برای بازشناسی استفاده کند. از آنجایی که مجموعه افراد مشابه روی یک رویه قرار می‌گیرند، با استفاده از الگوریتم‌های رتبه‌بندی رویه که با استفاده از گراف و همسایگی داده‌ها برچسب را در فضای گراف منتشر می‌کنند، می‌توان نزدیک‌ترین فرد را در مجموعه داده‌های موجود استخراج کرد. در واقع، داده‌های برچسب‌نخورده فضای رویه را پر کرده و در نتیجه، مزیت منتشر شدن برچسب از طریق آن‌ها را برای ما فراهم می‌آورند.

نسخه دیگری از این گروه در قالب مسئله رتبه‌بندی است. در مقاله [۲۶] فرض شده مجموعه آموزشی شامل تصاویر تهیه شده از یک فرد در دو دوربین مختلف، به صورت $X = \{(x_i, y_i)\}_{i=1}^m$ باشد که در آن بردار x_i بعدی ویژگی‌ها و y_i برچسب نمونه i می‌باشد. هدف در این رویکرد طراحی تابع رتبه‌بندی δ است؛ به طوری که برای نمونه‌های مرتبط مقدار بزرگ‌تری نسبت به نمونه‌های غیر مرتبط نتیجه دهد. تابع δ نیز به صورت زیر تعریف می‌شود که حاصل ضرب داخلی بردار وزن w در قدر مطلق بردار تفاضل x_i و $x_{i,j}$ است.

$$\delta(x_i, x_{i,j}^+) > \delta(x_i, x_{i,j}^-) \quad (13)$$

$$\delta(x_i, x_{i,j}) = w^T |x_i - x_{i,j}|$$

برقراری نامساوی ذکر شده در بالا منجر به حل مسئله‌ای مشابه ماشین بردار پشتیبان خواهد شد که در آن، بردار w برای همه نمونه‌های آموزشی طوری تعیین می‌شود که نامساوی بالا برقرار گردد. بدین ترتیب، برای یک زوج دوربین معیار فاصله‌ای آموزش می‌یابد که می‌تواند نگاهی از فاصله زوج تصاویر افراد ایجاد کند که در آن مقدار بیشتر، فرد مرتبط‌تری را پیشنهاد خواهد داد.

واقع، تبدیل فضایی از فضای ویژگی‌های اولیه حاصل می‌گردد که در این فضای جدید، فاصله نمونه‌های یک کلاس حتی المقدور کم و فاصله نمونه‌های کلاس‌های متفاوت زیاد باشد. به عبارتی، هدف یادگیری و آموزش این تبدیل است به نحوی که شرایط ذکر شده در بالا هرچه بیشتر برآورده شود. این یادگیری می‌تواند به صورت بانظرات صورت گیرد. در مرحله آموزش زوج داده‌هایی را که به یک فرد مربوط است، به عنوان داده‌های کلاس مثبت و زوج داده‌هایی را که به دو فرد متفاوت مربوط است، به عنوان نمونه کلاس منفی در نظر می‌گیریم. فرض کنید داده‌های مربوط به هر فرد با بردار m بعدی x نمایش داده شود. بدین ترتیب، هدف یادگیری معیار فاصله‌ای است که با ماتریس $D \in R^{m \times m}$ نمایش داده می‌شود. با استفاده از این ماتریس، فاصله دو نمونه x_i و x_j به صورت زیر محاسبه می‌شود [۱]:

$$d(x_i, x_j) = (x_i - x_j)^T D (x_i - x_j) \quad (13)$$

به کمک تعریف فاصله به صورت بالا هدف مینیمم کردن رابطه پایین به کمک برنامه‌نویسی تحدیبی است:

$$\min \sum_{(x_i, x_j) \in Pos} \|(x_i, x_j)\|_D^2$$

$$s. t. D \geq 0, \text{ and } \sum_{(x_i, x_j) \in Neg} \|(x_i, x_j)\|_D^2 \geq 1 \quad (14)$$

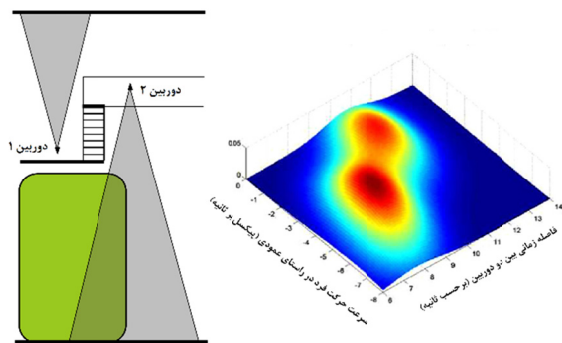
این رابطه چنین بیان می‌کند که می‌خواهیم ماتریس D را طوری به دست آوریم که مجموع فاصله زوج نمونه‌های مرتبط با یک فرد (Pos) مینیمم گردد و در عین حال، مجموع فاصله زوج نمونه‌های غیر مرتبط (Neg) بزرگ‌تر مساوی یک گردد.

نسخه‌ای از ایده مطرح شده در بالا در مقاله‌های [۲۳ و ۲۴] ارائه شده است. تفاوت نسخه ارائه شده در این مقالات، در این است که تنها شرطی که به دنبال برآورده نمودن آن هستیم، این است که احتمال اینکه فاصله نمونه‌های مرتبط کمتر از نمونه‌های غیر مرتبط باشد، حداکثر گردد، درحالی که در نسخه استاندارد،

• کالیبراسیون رنگ^۱

محیط‌هایی مانند بزرگراه، راهروهای اماکن و... تمایل عبور از مسیرهای مشترکی در طول زمان دارند. با استفاده از این فرض می‌توان فضای جست‌وجو را در مسئلهٔ بازشناسی محدود کرد. الگوریتم‌های ارائه‌شده در این دسته، سعی در یادگیری ارتباط بین دوربین‌ها در قالب تخمین یک تابع چگالی احتمال چند متغیره دارند. این متغیرهای تصادفی از جنس زمان و مکان‌اند.

در مقاله [۲۸] فرضیات مذکور را در قالب توابع توزیع احتمالاتی بیان کرده و در مرحلهٔ بعد، به کمک داده‌های آموزشی و با استفاده از روش پنجره پارزن^۳ توابع توزیع مذکور را تخمین زده است. این توابع توزیع احتمالاتی چندمتغیره، ارتباطات زیر را مدل می‌کنند: ارتباط بین زمان انتقال و اندازه و جهت حرکت اشیاء برای زوج دوربین‌ها، ارتباط بین زمان انتقال و موقعیت خروج از یک دوربین و موقعیت ورود به میدان دید دوربین دیگر، وابستگی بین موقعیت ورود و موقعیت خروج برای زوج دوربین‌ها.



شکل (۱۲): تابع توزیع تخمین شده سمت راست ارتباط بین سرعت خروج فرد از دوربین ۲ و زمانی که طول می‌کشد همان فرد به دوربین ۱ وارد شود (مستطیل سبز فضای سبزی را نشان می‌دهد که افراد کمتر وارد آن می‌شوند)

بدین ترتیب با تخمین این توابع توزیع، می‌توان فضای جست‌وجوی مسئله بازشناسی را محدود کرد و تصاویر کاندید کمتری را در نظر گرفت. شکل ۱۲ تابع توزیع احتمال دومتغیره برای ارتباط بین سرعت خروج فرد از دوربین ۲ و زمانی که طول می‌کشد همان فرد به دوربین ۱ وارد شود، نشان می‌دهد. نواحی قرمز رنگ در این نمودار، مناطقی هستند که احتمال وقوع

یکی از مشکلاتی که بر دقت بازشناسی تأثیر نامطلوب می‌گذارد، تغییرات نوری در دوربین‌های مختلف است. این تغییرات نوری می‌تواند میزان روشنایی یا حتی تبدیلات رنگی گوناگونی را شامل شود؛ برای مثال، ممکن است دوربینی رنگ سفید را با ته‌مایهٔ زرد و دوربین دیگری رنگ سفید را با ته‌مایهٔ آبی نشان دهد. روش‌هایی که در این گروه قرار می‌گیرند، سعی در تشخیص رابطهٔ رنگی بین دو دوربین دارند. به عبارتی هدف این روش، تعیین رابطهٔ نگاشت رنگی بین دو دوربین است. بنابراین، برای هر دو زوج دوربین این تابع و رابطه باید مورد یادگیری و به‌روزرسانی قرار گیرد. در مقالاتی که در این گروه قرار می‌گیرند [۲۷] از این تابع تحت عنوان BTF^2 یاد می‌شود.

۲.۲.۲. روش‌هایی که از اطلاعات پیکربندی و موقعیت

دوربین‌ها استفاده می‌کنند

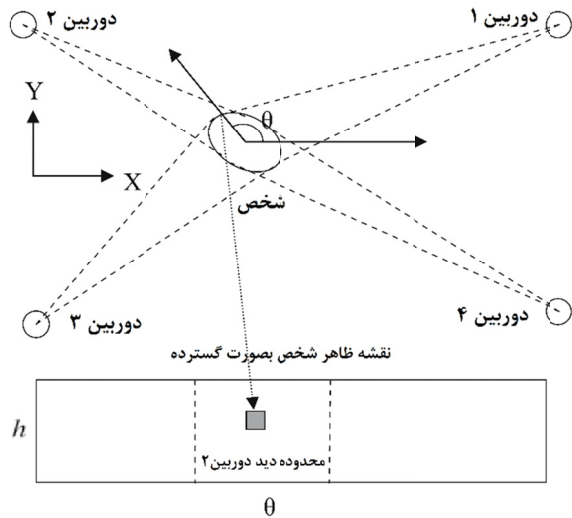
در این دسته از روش‌ها سعی داریم با استفاده از منابع اطلاعاتی غیر از ظاهر افراد، فضای جست‌وجو را محدود کنیم. منظور از فضای جست‌وجو افرادی است که در مجموعه گالری وجود دارند و در نتیجه، هر یک کاندیدی برای بازشناسی‌اند؛ برای مثال، با در دست داشتن ارتباط مکانی دوربین‌های نصب‌شده می‌توان حدس زد که براساس یک توزیع احتمالاتی فرد خاصی که به میدان دید یک دوربین وارد شده، احتمالاً در لحظاتی قبل در کدام دوربین‌ها می‌توانسته ظاهر گردد. بدین ترتیب، تعداد افراد کاندید با توجه به قید جدید، محدودتر شده و دقت بازشناسی افزوده می‌شود.

یکی از مواردی که منجر به پیچیدگی بازشناسی می‌شود، وجود دوربین‌های متعدد با دامنهٔ دید ناهم‌پوشان است. وجود چنین شرایطی گستردگی مشاهدات را در زمان و مکان‌های گوناگون در دوربین‌های مختلف دربرخواهد داشت. قاعده‌ای که به‌طور معمول می‌توان فرض کرد این است که اشیاء متحرک در

1. Color Calibration

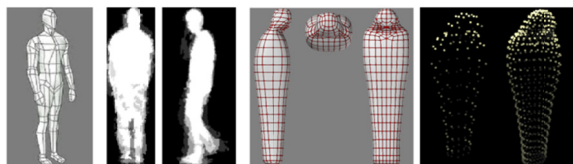
2. Brightness Transfer Function

3. Parzen window



شکل (۱۳): کالیبراسیون چند دوربین با میدان دید مشترک برای یک فرد

یکی دیگر از مقالات ارائه شده در این دسته، تحقیقات بالتیری و همکاران است که با استفاده از اطلاعات دوربین‌های کالیبره شده، اقدام به تشکیل مدلی سه بعدی از انسان می‌کنند [۳۰]. پس از تشخیص موقعیت افراد در دوربین‌ها با توجه به مسیر حرکت افراد جهت شخص، اندازه و موقعیت هریک مشخص می‌شود. یکی از ایده‌های مطرح شده در این مقاله، ارائه مدلی سه بعدی و ساده شده از پیکر انسان است که در عین سادگی، بسیاری از خصوصیات سه بعدی انسان را دارد. شکل ۱۴ مدل پیشنهادی را نشان می‌دهد.



شکل (۱۴): الف. مدل سه بعدی انسان را با جزئیات نشان می‌دهد؛ ب. میانگین تصویر پیش‌زمینه انسان را نشان می‌دهد؛ پ. مدل سه بعدی ساده شده پیشنهادی را نشان می‌دهد؛ ت. نمونه برداری از رئوس مورد نیاز از مدل ساده سازی شده را نشان می‌دهد.

بیشتری دارد. این نواحی در این نمودار نشان می‌دهد که هرچه سرعت خروج فرد از دوربین ۲ بیشتر باشد، زمانی که طول می‌کشد که در دوربین ۱ ظاهر شود، کمتر خواهد بود یا به عبارتی برای سرعت‌های زیاد و زمان انتقال کم، تابع چگالی مقدار بیشتری را نشان می‌دهد.

• کالیبره کردن دوربین‌ها

در این دسته از روش‌ها معمولاً دوربین‌های نصب شده در محیط دارای میدان دید هم‌پوشان هستند. بدین ترتیب می‌توان هم‌نگاری^۱ بین دوربین‌ها را استخراج کرد. منظور از هم‌نگاری استخراج نگاشتی است که تحت نگاشت مذکور یک صفحه رؤیت شده در یک دوربین به تصویر همان صفحه در دوربین دوم تبدیل می‌گردد. از این عمل تحت عنوان کالیبراسیون نیز یاد می‌شود. به عبارتی، دوربین‌های مختلف در این فرایند هم‌تراز می‌شوند. پس از هم‌ترازی دوربین‌ها می‌توان خصوصیات یکتایی از اشیاء رؤیت شده استخراج کرد. این خصوصیات می‌تواند موقعیت سه بعدی شیء در محیط، مدل سه بعدی شیء، ارتفاع شیء و... باشد.

در مقاله [۲۹] با در دست داشتن دنباله‌ای از تصاویر فرد از چند زاویه دید، مدلی سراسرنما^۲ ساخته می‌شود. به عبارتی، نقشه سراسرنما براساس اطلاعات به دست آمده از دوربین‌های یکه دارای هم‌پوشانی هستند، به عنوان ویژگی فرد در نظر گرفته شده است. در نهایت، برای مقایسه تصویر جدید، مقایسه روی هیستوگرام مدل و هیستوگرام مورد آزمون صورت می‌گیرد. شکل ۱۳ مثالی از شخصی است که بین چهار دوربین واقع شده و بدین ترتیب، دوربین‌ها با هم کالیبره شده و زوایای تصویربرداری به دست آمده و در نهایت، تصاویر روی یک مدل سراسرنمای استوانه‌ای ادغام می‌شوند.

1. Homography

2. Panoramic

سپس در خوشه مذکور با استفاده از بردار وزنی متناظر با ویژگی‌های آن، تصاویر موجود در خوشه مذکور رتبه‌بندی می‌شوند.

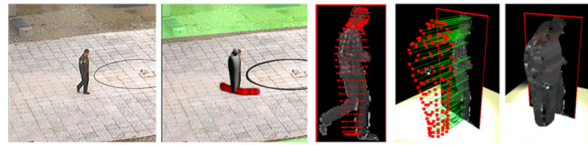
در مقاله [۳۲] با در نظر گرفتن ویژگی‌های سطح میانی که برای انسان نیز قابل فهم می‌باشد، کیفیت بازشناسی را بهبود داده است. بدین ترتیب که ابتدا ۱۲ ویژگی توسط یک خبره تعیین شده و سپس با استفاده از یک پایگاه داده برای هر ویژگی یک کلاس بند طراحی می‌گردد. برخی از ویژگی‌های مذکور عبارت‌اند از شلوار کوتاه، دامن پوش، موبلند، موکوتاه، قدبلند، قدکوتاه، حمل‌کننده یک شیء، کوله‌پشتی دار و ... سپس برای بازشناسی یک فرد جدید با استفاده از هریک از این کلاس‌بندها مشخص می‌شود که فرد جدید کدام‌یک از خصیصه‌های مذکور را دارد. بدین ترتیب، برای هر فرد، برداری از وجود یا عدم وجود خصیصه‌ها را خواهیم داشت که با ترکیب برداری از مقادیر این کلاس‌بندها و روش‌های گذشته، کیفیت بازشناسی بهبود می‌یابد.

در مقالات [۲۳ و ۳۳] سعی بر استخراج نواحی بااهمیت برای یک فرد است. روش کار بدین صورت است که با استفاده از مفهوم انطباق متراکم، به کمک توصیف‌گر SIFT^۱ ابتدا با مقایسه تصویر با مجموعه تصادفی از تصاویر انسان تکه‌هایی را که در بیشتر از نیمی از افراد دیده نشده، به‌عنوان نواحی بااهمیت مشخص کرده است و در نهایت، در رابطه‌ای ترکیبی و وزن‌دار، این نواحی بااهمیت را وزن‌دهی بیشتری کرده و نشان می‌دهد که با این روش نرخ تشخیص بالاتر رفته است.

۳. معیارهای ارزیابی بازشناسی انسان

معیارهای ارزیابی بازشناسی را می‌توان در دو حالت مجموعه بسته و مجموعه باز بررسی کرد. در حالت مجموعه بسته که مجموعه موارد آزمون در مجموعه گالری وجود دارد، مسئله بازشناسی در قالب یک مسئله رتبه‌بندی در نظر گرفته می‌شود. در این حالت، به‌منظور نشان دادن دقت تشخیص از نمودار

پس از تعیین جهت شخص، مدل سه‌بعدی روی آن قرار گرفته و سپس پیکسل‌های تصویر فرد در جهت مناسب روی رئوس مدل، تصویر و انعکاس داده می‌شود. مراحل این فرایند در شکل ۱۵ نشان داده شده است.



شکل (۱۵): فرایند انطباق تصویر فرد روی مدل سه‌بعدی پیشنهادی

از ویژگی‌های الگوریتم پیشنهادی در مقاله مذکور این است که اطلاعات به‌دست‌آمده از یک شخص خاص که به‌مرور زمان مشخص می‌شود، منجر به به‌روزرسانی مدل می‌گردد. این به‌روزرسانی می‌تواند منجر به تکمیل مدل از زوایای مختلف گردد یا منجر به وضوح بیشتر در ناحیه خاصی از مدل شود.

۳.۲.۲. وزن‌دهی و اهمیت‌دهی به ویژگی‌ها

در این دسته به کمک وزن‌دهی و اهمیت‌گذاری روی ویژگی‌ها، سعی بر محدود کردن فضای جست‌وجو و در واقع اعمال نگاشتی است که نتیجه جست‌وجو حاوی جواب‌های مرتبط‌تری باشد. در مقاله [۳۱] سعی در وزن‌دهی و اهمیت‌دهی در ویژگی‌های سطح پایین با توجه به خصیصه‌های استخراج‌شده دارد. مراحل انجام الگوریتم ارائه‌شده بدین ترتیب است که ابتدا تصاویر برچسب‌نخورده را خوشه‌بندی می‌کند که بدین ترتیب خروجی خوشه‌بندی مجموعه‌ای از نماینده‌ها خواهد بود که هرکدام بازگوکننده خصیصه‌ای می‌باشند.

روش خوشه‌بندی صورت‌گرفته با استفاده از پارتیشن‌بندی گراف است. در قدم بعدی الگوریتم باید برای هر نماینده یک بردار وزن متناظر با ویژگی‌های سطح پایین محاسبه شود. وزن هر ویژگی بدین صورت محاسبه می‌شود که به کمک یک ویژگی داده‌های نماینده‌های مختلف را دسته‌بندی می‌کنیم. هرچه یک ویژگی بهتر بتواند یک نماینده خاص را از دیگر نماینده‌ها مجزا کند، برای نماینده مذکور از اهمیت بالاتری برخوردار است. در مرحله بازشناسی، زمانی که تصویر یک فرد جدید وارد سیستم می‌شود، ابتدا خوشه یا نماینده تصویر ورودی محاسبه می‌شود.

1. Scale-invariant feature transform

در این رابطه، MMs تعداد مواردی را نشان می‌دهد که مورد آزمون در مجموعه گالری بوده، ولی نادرست تشخیص داده می‌شود و FPS تعداد مواردی را نشان می‌دهد که مورد آزمون در مجموعه گالری نیست، اما الگوریتم آن را برچسب‌گذاری کرده است. به‌منظور مشخص کردن رفتار سیستم نمودار Accuracy-FAR رسم می‌شود؛ برای مثال، می‌توان مقدار آستانه‌ای را که سیستم تصمیم به رد کردن یک نمونه مورد آزمون می‌کند، از ۰ تا ۱ با گام‌های ۰.۰۵ مقدار دقت و نرخ پذیرش خطا را محاسبه کرد و همهٔ این اطلاعات را بر روی نمودار Accuracy-FAR مشاهده کرد.

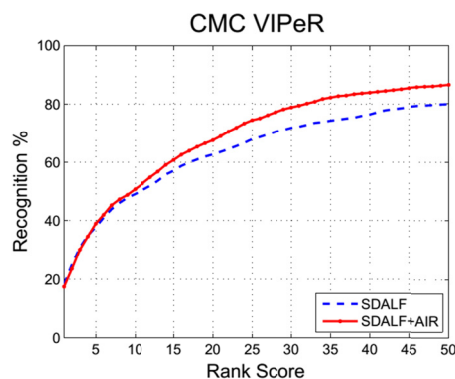
۴. مجموعه داده‌ها برای ارزیابی عملکرد سیستم نظارت ویدئویی

به‌منظور ارزیابی الگوریتم‌های بازشناسی مجموعه داده‌های متنوعی در دسترس است که هرکدام چالش‌های خاصی را در مسئلهٔ بازشناسی مطرح می‌کنند. تعدادی از این مجموعه داده‌ای عبارت‌اند از:

ViPER: این مجموعه داده‌ای شامل تصویر ۶۳۲ فرد است. برای هر فرد دو تصویر مربوط به دو دوربین مختلف تهیه شده است. دو تصویر مربوط به یک فرد از زوایای مختلف‌اند. شرایط نوری در دو دوربین متنوع است. تصاویر هر فرد به مستطیلی محدود شده و در اندازهٔ ۱۲۸ در ۴۸ پیکسل می‌باشد. این مجموعه داده سخت‌ترین مجموعه داده‌ای است که در بازشناسی مورد استفاده قرار گرفته است [۳۴].

ETHZ: این مجموعه داده‌ای شامل تصاویر گرفته‌شده به‌وسیلهٔ یک دوربین متحرک است. از هر شخص چندین تصویر موجود است؛ از هر شخص سه دنباله تصاویر که به ترتیب شامل ۸۳، ۳۵ و ۲۸ تصویر در اندازه‌های مختلف و تصاویر شامل هم‌پوشانی نیز می‌باشد [۳۵].

CMC^۱ استفاده می‌شود. در این نمودار، روی محور افقی رتبه‌های مختلف را رسم کرده است؛ برای مثال، نمودار شکل ۱۶ نشان می‌دهد زمانی که الگوریتم SDALF نمونه‌های گالری را برای تمام حالات مورد آزمون مرتب می‌کند، به‌طور میانگین، در ۲۰ درصد موارد نمونه درست در رتبهٔ ۱ قرار می‌گیرد. یا برای مثال نشان می‌دهد الگوریتم SDALF در ۸۰ درصد موارد نمونه درست در رتبه‌ای زیر ۵۰ قرار می‌گیرد.



شکل (۱۶): یک نمونه از نمودار CMC

زمانی که مسئله به‌صورت مجموعه باز است، نمودار CMC نمی‌تواند به‌درستی تمام ویژگی‌های الگوریتم را نشان دهد. در حالت مجموعه باز، مواردی خواهیم داشت که نمونه مورد آزمون در مجموعه گالری نیست. در این حالت، ارزیابی باید براساس دو معیار دقت (رتبهٔ ۱) و نرخ پذیرش خطا محاسبه (FAR^۲) گردد. منظور از دقت همان موارد تشخیص TP^۳ می‌باشد که نشان‌دهندهٔ درصد نمونه‌هایی از مجموعه آزمون است که به‌درستی تشخیص داده شده است. رابطهٔ نرخ پذیرش خطا به‌صورت زیر است:

$$FAR = \frac{MMs + FPS}{N_p} \quad (16)$$

1. Cumulative Matching Characteristic
2. False Acceptance Rate
3. True Positive

جدول (۱): مقایسه برخی از مهم‌ترین روش‌های موجود در بازشناسی انسان

روش‌های ارائه‌شده	توصیف‌گرهای مورد استفاده	نوع ویژگی مورد استفاده	استفاده از مدل انسان	داده‌های آزمایشی / درصد تشخیص CMC (Rank 1)
[۳]		بافت رنگ	بله	(~ 60%) اختصاصی
[۴]	AHPE	بافت رنگ	تا حدی	ETHZ1 (~ 90%) ETHZ2 (~ 90%) ETHZ3 (~ 95%) iLIDS (~ 35%) CAVIAR4REID (~ 10%)
[۶]	MSCR RHSP	بافت رنگ	تا حدی	ETHZ1 (~ 90%) ETHZ2 (~ 90%) ETHZ3 (~ 95%) CAVIAR4REID (~ 8%) ViPER (~ 19%)
[۷]	CovDesc	بافت	خیر	(~ 64%) اختصاصی
[۸]	CovDesc Gabor	بافت رنگ	خیر	ViPER (~ 20%) ETHZ1 (~ 93%) ETHZ2 (~ 91%) ETHZ3 (~ 96%)
[۹]	CovDesc	بافت رنگ	خیر	iLIDS-AA (~ 78%)
[۲۰]	CovDesc HOG	بافت	تا حدی	iLIDS (~ 34%)
[۱۲]	CovDesc	بافت	خیر	iLIDS-AA (~ 44%) ETHZ (~ 97%)
[۱۳]	Probilistic Histogram	رنگ	خیر	ViPER (~ 7%)
[۱۴]	Fuzzy Histogram	رنگ	خیر	ViPER (~ 12%)
[۳۸]	Gabor, Schmid Spatiogram	بافت رنگ	تا حدی	iLIDS (~ 20%) ViPER (~ 10%)
[۱۷]	Gabor Histogram Schmid	بافت رنگ	خیر	ViPER (~ 12%)
[۱۸]		رنگ لبه	خیر	ViPER (~ 19%)
[۲۱]		بافت رنگ	خیر	ViPER (~ 14%) CAVIAR4REID (~ 29%)
[۲۲]		بافت رنگ	خیر	ViPER (~ 30%) CUHK (~ 28%)
[۲۳]			خیر	ViPER (~ 15%) ETHZ (~ 72%) iLIDS (~ 44%)
[۲۴]			خیر	iLIDS (~ 49%) ViPER (~ 19%) GRID (~ 12%)
[۲۵]	Gabor	بافت رنگ	خیر	iLIDS (~ 31%) ViPER (~ 14%)
[۳۰]	Gabor	بافت رنگ، لبه	خیر	iLIDS (~ 35%) ViPER (~ 16%)
[۳۲]	SIFT Histogram	بافت رنگ	خیر	ViPER (~ 27%)

iLIDS: این مجموعه داده‌ای شامل ۴۷۶ تصویر از ۱۱۹ فرد می‌باشد که از دو دوربین با میدان دید ناهم‌پوشان به دست آمده است. به‌طور میانگین، برای هر فرد ۴ تصویر موجود است. تصاویر شامل شرایط متفاوت نوری و شامل هم‌پوشانی نیز می‌باشند. تصاویر در اندازه ۱۲۸ در ۶۴ تهیه شده است [۳۶].

CAVIAR4REID: این مجموعه داده‌ای شامل تصاویری از دو دوربین با میدان دید مشترک می‌باشد که از محیط یک فروشگاه تصویربرداری شده است. تصاویر موجود مربوط به ۷۲ فرد مختلف است که انواع شرایط مشکل‌ساز از قبیل شرایط نوری متفاوت، حالت و پوزیشن و هم‌پوشانی را شامل می‌شود [۳۷].

V-47: این مجموعه داده‌ای شامل ویدئوهایی از ۴۷ شخص از دو دوربین مختلف در فضای داخلی گرفته شده است. برای هر دوربین، دو ویدئو در دو زاویه تهیه شده که این دو زاویه، دو جهت راه رفتن افراد می‌باشد. تغییرات نوری و اندازه و شلوغی در این مجموعه داده‌ای زیاد نیست [۳۸].

۵. بحث و نتیجه‌گیری

جدول ۱ مقایسه‌ای از مهم‌ترین روش‌های ارائه‌شده تاکنون را نشان می‌دهد. همان‌طور که در این جدول مشاهده می‌کنید، توصیف‌گرهای گوناگونی به‌منظور بازنمایی انسان مورد استفاده قرار گرفته است. دو نکته در این جدول قابل تأمل است. اولین نکته دقت پایین روش‌ها در بازشناسی مجموعه داده‌های ViPER است. با بررسی ویژگی‌های این مجموعه داده‌ای درمی‌یابیم که تصاویر افراد در این مجموعه از زوایای مختلف می‌باشد.

روش‌ها پیکره انسان به دو بخش بالانه و پایین‌تنه تقسیم می‌شود. این تقسیم‌بندی پیکره، استفاده از این دانش پیشین است؛ لذا روند تحقیقات می‌تواند به سمت ارائه و طراحی مدل‌های کامل‌تری از انسان باشد که نسبت به انواع مشکلات و خصوصاً مشکل تغییرات زاویه مقاوم عمل کنند.

فرض گوناگونی زوایای دید فرضی معقول بوده که در سیستم‌های واقعی نیز با چنین شرایطی مواجهیم. نکته دوم بحث استفاده از ویژگی‌های هندسی پیکره انسان است. به عبارتی، روش‌هایی که از این دانش پیشین (ویژگی‌های هندسی مشترک بین نمونه‌های انسان) بهره می‌برند، دقت تشخیص بالاتری را عرضه می‌دارند. برای مثال در گروهی از

مراجع

- [1] Bedagkar-Gala, A., Shah, S. K., "A survey of approaches and trends in person re-identification", Image and Vision Computing, Elsevier, Vol. 32, pp. 270-286, 2014.
- [2] Vezzani, R., Baltieri, D., and Cucchiara, R., "People reidentification in surveillance and forensics: A survey", ACM Computing Surveys (CSUR), Vol. 46, pp. 1-29, 2013.
- [3] Gheissari, N., Sebastian, T. B., and Hartley, R., "Person re-identification using spatiotemporal appearance", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1528-1535, 2006.
- [4] Bazzani, L., Cristani, M., Perina, A., and Murino, V., "Multiple-shot person re-identification by chromatic and epitomic analyses", Pattern Recognition Letters, Elsevier, Vol. 33, pp. 898-903, 2012.
- [5] Jovic, N., Perina, A., Cristani, M., Murino, V., and Frey, B., "Stel component analysis: Modeling spatial correlations in image class structure", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2044-2051, 2009.
- [6] Bazzani, L., Cristani, M., and Murino, V., "Symmetry-driven accumulation of local features for human characterization and re-identification" Computer Vision and Image Understanding, Elsevier, Vol. 117, pp. 130-144, 2013.
- [7] Metzler, J., "Appearance-Based Re-identification of Humans in Low-Resolution Videos Using Means of Covariance Descriptors", The 9th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), pp. 191-196, 2012.
- [8] Ma, B., Su, Y., and Jurie, F., "Bicov: a novel image representation for person re-identification and face verification", British Machine Vision Conference, 2012.
- [9] Bąk, S., Charpiat, G., Corvée, E., Brémond, F., and Thonnat, M., "Learning to match appearances by correlations in a covariance metric space", The 12th European Conference on Computer Vision, pp. 806-820, 2012.
- [10] Zhang, Y. and Li, S., "Gabor-LBP based region covariance descriptor for person re-identification", The 6th International Conference on Image and Graphics (ICIG), pp. 368-371, 2011.
- [11] Bak, S., Corvee, E., Brémond, F., and Thonnat, M., "Person re-identification using spatial covariance regions of human body parts", Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 435-440, 2010.
- [12] Bak, S., Corvee, E., Brémond, F., and Thonnat, M., "Multiple-shot human re-identification by mean riemannian covariance grid", The 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), pp. 179-184, 2011.
- [13] D'Angelo, A. and Dugelay, J.-L., "People re-identification in camera networks based on probabilistic color histograms", The Electronic Imaging Conference on 3D Image Processing and Applications, 2011.

- [14] Xiang, Z. J., Chen, Q., and Liu, Y., "*Person re-identification by fuzzy space color histogram*", Multimedia Tools and Applications, Springer, Vol. 73, No. 1, pp. 91-107, 2012.
- [15] Birchfield, S. T., and Rangarajan, S., "*Spatiograms versus histograms for region-based tracking*", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1158-1163, 2005.
- [16] Satta, R., Fumera, G., and Roli, F., "Fast person-identification based on dissimilarity representations", Elsevier, Pattern Recognition Letters, Vol. 33, pp. 1838-1848, 2012.
- [17] Khan, A., Zhang, J., and Wang, Y., "*Appearance-based re-identification of people in video*", International Conference on Digital Image Computing: Techniques and Applications, pp. 357-362, 2010.
- [18] Gray, D. and Tao, H., "*Viewpoint invariant pedestrian recognition with an ensemble of localized features*", The 10th European Conference on Computer Vision, pp. 262-275, 2008.
- [19] Hirzer, M., Beleznai, C., Roth, P. M., and Bischof, H., "*Person re-identification by descriptive and discriminative classification*", The 17th Scandinavian Conference on Image Analysis, pp. 91-102, 2011.
- [20] Bąk, S., Corvee, E., Bremond, F., and Thonnat, M., "*Boosted human re-identification using riemannian manifolds*", Image and Vision Computing, Elsevier, Vol. 30, pp. 443-452, 2012.
- [21] Bąk, S., Corvee, E., Bremond, F., and Thonnat, M., "*Person re-identification using haar-based and dcd-based signature*", The 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 1-8, 2010.
- [22] Avraham, T., Gurvich, I., Lindenbaum, M., and Markovitch, S., "*Learning implicit transfer for person re-identification*", European Conference on Computer Vision (ECCV), pp. 381-390, 2012.
- [23] Zhao, R., Ouyang, W., and Wang, X., "*Person re-identification by salience matching*", IEEE International Conference on Computer Vision, pp. 2528-2535, 2013.
- [24] Zheng, W.-S., Gong, S., and Xiang, T., "*Reidentification by relative distance comparison*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, pp. 653-668, 2013.
- [25] Loy, C. C., Liu, C., and Gong, S., "*Person re-identification by manifold ranking*", IEEE International Conference on Image Processing, pp. 3567-3571, 2013.
- [26] Prosser, B., Zheng, W. S., Gong, S., Xiang, T., and Mary, Q., "*Person Re-Identification by Support Vector Ranking*", The British Machine Vision Conference, pp. 21.1-21.11, 2010.
- [27] Javed, O., Shafique, K., and Shah, M., "*Appearance modeling for tracking in multiple non-overlapping cameras*", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 26-33, 2005.
- [28] Javed, O., Shafique, K., Rasheed, Z., and Shah, M., "*Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views*", Computer Vision and Image Understanding, Elsevier, Vol. 109, pp. 146-162, 2008.
- [29] Gandhi, T., Trivedi M. M., "*Person tracking and reidentification: Introducing Panoramic Appearance Map (PAM) for feature representation*", Machine Vision and Applications, Springer, Vol. 18, pp. 207-220, 2007.
- [30] Baltieri, D., Vezzani, R., and Cucchiara, R., "*Sarc3d: a new 3d body model for people tracking and re-identification*", The International Conference on Image Analysis and Processing, pp. 197-206, 2011.
- [31] Liu, C., Gong, S., Loy, C. C., and Lin, X., "*Person re-identification: what features are important?*". European Conference on Computer Vision, pp. 391-401, 2012.
- [32] Layne, R., Hospedales, T. M., and Gong, S., "*Towards person identification and re-identification with attributes*", European Conference on Computer Vision, pp. 402-412, 2012.
- [33] Zhao, R., Ouyang, W., and Wang, X., "*Unsupervised salience learning for person re-identification*", IEEE Conference on Computer

- Vision and Pattern Recognition, pp. 3586-3593, 2013.
- [34] Gray, D., Brennan, S., and Tao, H., "*Evaluating appearance models for recognition, reacquisition, and tracking*", IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 2007.
- [35] Ess, A., Leibe, B., and Van Gool, L., "*Depth and appearance for mobile scene analysis*" The 11th IEEE International Conference on Computer Vision, pp. 1-8, 2007.
- [36] U.H. Office, i-Lids Multiple Camera Tracking Scenario Definition, Available: www.homeoffice.gov.uk/science-research/hosdb/i-lids/2008.
- [37] <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/2004>.
- [38] Wang, S., Lewandowski, M., Annesley, J., and Orwell, J., "*Re-identification of pedestrians with variable occlusion and scale*", IEEE International Conference on Computer Vision, pp. 1876-1882, 2011.